Forecasting Shark Attack Risk Using AI: A Deep Learning Approach

Evan Valenti

SafeWaters.AI, Boston, USA
Email: evan@safewaters.ai

Abstract

This study aimed to develop a predictive model utilizing available data to forecast the risk of future shark attacks, making this critical information accessible for everyday public use. Employing a deep learning/neural network methodology, the system was designed to produce a binary output that is subsequently classified into categories of low, medium, or high risk. A significant challenge encountered during the study was the identification and procurement of appropriate historical and forecasted marine weather data, which is integral to the model's accuracy. Despite these challenges, the results of the study were startlingly optimistic, showcasing the model's ability to predict with impressive accuracy. In conclusion, the developed forecasting tool not only offers promise in its immediate application but also sets a robust precedent for the adoption and adaptation of similar predictive systems in various analogous use cases in the marine environment and beyond.

1. Introduction

1.1. The Importance of Forecasting Shark Attacks

The waters surrounding our coastlines offer innumerable opportunities for recreation, livelihood, and exploration. Yet, as is the nature of the vast marine environment, it also holds certain risks. Foremost among these for many beach-goers and marine enthusiasts is the potential for shark attacks. Safeguarding these waters, SafeWaters has taken the mantle to mitigate shark attack risks for the American populace. A glaring statistic to note is that in 2021, 47 Americans

became victims of unprovoked shark attacks, accounting for a staggering 64% of the worldwide total of such incidents. This translates to 137 recorded shark bites across the globe, of which 11 resulted in fatalities [1]. Such numbers are not merely anomalous blips on the radar. Historical data reveals an escalating trend, with the number of unprovoked shark attacks in the U.S. escalating from a mere four in 1955 to a concerning 57 in 2015 [2]. These statistics underscore the pressing need for robust measures to predict and prevent such unfortunate encounters.

1.2. Current Methods and Their Limitations

Present-day strategies to mitigate shark attacks, while well-intentioned, often introduce more problems than they solve. A prime example is the utilization of shark nets. While designed to protect swimmers and surfers, these nets have a devastating environmental impact. Alarmingly, shark nets contribute to the death of millions of sharks every month. To put this into perspective, approximately 273 million sharks meet untimely deaths annually due to such human interventions [3]. These measures not only diminish the shark population but also disrupt marine ecosystems, where sharks play a pivotal role as apex predators. The repercussions of these actions resonate through the food chain, leading to imbalances that may, in the long run, prove even more detrimental to marine life and human activity in coastal areas. Yet, these aren't the only shortcomings. Many of the prevalent methods are reactive, localized, and fail to provide real-time or future-oriented insights.

1.3. Why an AI Approach Might Offer Advantages

In the ever-evolving landscape of technology, Artificial Intelligence (AI) stands out as a beacon of potential, particularly in the domain of predictive analytics. Utilizing past data, AI models can discern patterns and trends that are often too intricate for human analysis. This retrospective analysis is pivotal, but AI's real prowess lies in its ability to forecast future events based on these patterns. By implementing an AI-based approach to shark attack prediction, there's the potential to provide timely warnings, allowing beachgoers and authorities to take early precautions. Furthermore, such a tool can be constantly updated with real-time data, ensuring that its predictions are always grounded in the latest available information. Beyond just shark attacks, the adaptability of AI means that similar models can be employed for a myriad of other marine-related predictions, ushering in a new era of maritime safety and ecological awareness.

2. Methods

2.1. Data Collection

2.1.1. Global Shark Attack File Utilization
To establish the foundation for the study, we tapped into the comprehensive resource of the Global Shark Attack File. This repository offers a holistic dataset, capturing various instances of shark attacks, encompassing both provoked and

unprovoked encounters. The primary focus was to extract information relevant to the parameters of our study.

2.1.2. Dataset Cleaning and Structuring

Upon downloading the dataset, our first objective was to curate the information to ensure clarity and relevance. All rows, save for those representing the 'date' and 'location' of the attacks, were pruned. To facilitate the machine learning process, a new column titled "attack" was appended to the dataset. In this column, a "1" was placed against each entry, serving as a binary indicator of the occurrence of a shark attack on that specific date and location.

2.1.3. Geocoding the Dataset

For the next phase, we employed Python, a versatile programming language renowned for its data manipulation capabilities. With a script crafted in Python, the dataset was looped to geocode each location entry. This was accomplished by leveraging the robust capabilities of Google's Geocode API, which returned latitude and longitude coordinates for each listed location. Consequently, these coordinates were appended to the dataset, associating each shark attack event with its precise geographical point.

2.1.4. Integrating Marine Weather Data

Recognizing the potential correlation between marine weather conditions and shark behavior, it was deemed essential to incorporate marine weather data. A Historic Marine Weather API was utilized to fetch relevant weather data for each row indicating an attack. The fetched data was subsequently output to a new CSV file.

2.1.5. Expanding the Dataset

To ensure a thorough representation of both shark attack and non-attack days, another Python script was conceived. This script was responsible for retrieving marine weather data for every day of the year, for each beach globally with a documented shark attack. This was traced back until 2015, governed by the constraints of the available historical data.

2.1.6. Final Dataset Compilation

Upon retrieval, individual CSVs were concatenated, culminating in a comprehensive dataset. For each location, days devoid of any recorded shark attacks were introduced into the dataset. These entries were labeled with a "0" in the "attack" column, offering a binary distinction for machine learning algorithms to discern between days with and without shark attacks. The expansive dataset is full with detailed insights and granular data points that encompass the entire spectrum of marine conditions and their potential influences on shark behavior. The masking + indicator column method was applied to handle any missing data.

Through this meticulous process, we ensured that our dataset not only captured instances of shark attacks but also provided a holistic view of marine con-

ditions, geographical coordinates, and the relative frequency of these events. This expansive dataset served as the cornerstone for our predictive modeling. The collected data is most appropriate for this use case as marine weather conditions have been the backbone for understanding attacks throughout attack research history, already knowing cloudy days & murky waters have been the conditions for a magnitude of attacks. Expanding the scope of marine weather variable an letting our advanced neural network learn the relationships and weights of the 30 different monitored marine weather conditions is essential to learn their impact on the sharks behavior and aggression.

2.2. Model Selection

2.2.1. Selection of a Fully Connected Neural Network with Binary Output

In the realms of Artificial Intelligence and machine learning, multiple models exist, each with its specific capabilities, advantages, and potential shortcomings. For the task at hand—predicting the risk of shark attacks based on historical and meteorological data—a fully connected neural network (FCNN) was selected.

2.2.2. Rationale behind Choosing the Fully Connected Neural Network

The FCNN was considered apt for several reasons:

1) **Comprehensive Feature Learning**: FCNNs have the ability to automatically and adaptively learn spatial hierarchies of features from input data. Given the multidimensional nature of our dataset—encompassing geographical coordinates, marine weather conditions, and temporal data—the ability of the FCNN to capture intricate patterns in such datasets made it a logical choice.

2) **Binary Classification**: Our objective was to predict the occurrence or non-occurrence of a shark attack, which is essentially a binary classification problem. FCNNs, when combined with a sigmoid activation function in the output layer, are adept at such binary classifications.

3) **Flexibility**: The neural network architecture allows for easy adjustments. By tweaking the number of layers and nodes, or neurons, in each layer, the network can be adapted to handle varying complexities in data.

2.2.3. Training, Validation, and Testing Procedures

- **Data Preprocessing**:
- Date and time were parsed from strings to datetime objects, enabling the extraction of meaningful features like "Month", "Day", "Hour", and "Minute".
- Categorical features like "moon_phase", "weatherDesc", and "swellDir16Point" were transformed using one-hot encoding to convert them into a machine-readable format without introducing ordinal relationships where none exist.
- The MinMaxScaler was applied to normalize the features to ensure that no variable overshadows another due to differences in their magnitudes.
- **Dataset Splitting**:
- The data was divided into training and testing sets (70% and 30%, respectively) using stratified sampling, ensuring that the proportion of positive (attack) and

negative (no attack) samples remained consistent across both sets.

- **Neural Network Architecture**:
- The network comprises an input layer, three hidden layers, and an output layer.
- To prevent over fitting and improve generalization, dropout layers were introduced between the hidden layers.
- The final layer employed a sigmoid activation function, aligning with our binary classification objective.
- **Model Compilation and Training**:
- The model was compiled using the Adam optimizer and binary cross-entropy as the loss function—standard for binary classification tasks.
- Early stopping was incorporated into the training process, monitoring the validation loss. This halts training if the model's performance on the validation data does not improve after ten epochs, ensuring efficient training and preventing over fitting.
- The model was then trained using the training set, setting aside 20% of it for validation.
- **Model Evaluation**:
- Once trained, the model's predictions on the test set were converted to binary values (1 for "attack" and 0 for "no attack").
- Evaluation metrics like precision, recall, and F1-score were calculated for the positive class (attack instances) to gauge the model's accuracy in predicting actual shark attacks.

2.2.4. Conclusion

The chosen FCNN was specifically designed, structured, and optimized for the nature of the data at hand and the binary classification objective. With its deep layers and data preprocessing steps, the model efficiently learned from the historical data to predict shark attack occurrences, positioning it as a promising tool for safety measures and decision-making in marine activities.

3. Implementation

3.1. Introduction to the Implemented Technology

To make forecasting shark attacks as accessible and user-friendly as possible, the AI-driven solution was integrated into a mobile application. This allows users to quickly and seamlessly determine the risk of shark attacks for their chosen location, backed by a powerful AI model trained on historical and marine weather data.

3.2. Functionality of the Mobile App

- **User Interface**:
- Upon launching the app, users are greeted with a simple interface prompting them to input their location of interest.
- **Processing the User's Input**:

- Once the location is submitted, the mobile app sends this data as a POST request to an endpoint. This endpoint is hosted on a Flask application, which is set up and running on Google Cloud, ensuring high availability and scalability.
- **Geocoding the Location**:
- The Flask application processes the incoming location data and uses a geocoding service to convert the provided location name into its corresponding latitude and longitude coordinates. This is a vital step, as our AI model and marine weather API both require specific geographical coordinates for precise forecasting.
- **Fetching Marine Weather Forecasts**:
- With the derived latitude and longitude in hand, the Flask app then sends a request to the marine weather API. In response, the API provides marine weather forecasts for the upcoming seven days for the specified location.
- **Running the AI Model for Predictions**:
- The fetched seven-day marine weather forecast is input into the h5 file containing the weights of the trained model. This model processes the weather data and returns a seven-day risk prediction in the form of binary outputs (0 or 1 for each day).
- **Classifying and Displaying the Risk**:
- The binary outputs from the model are classified into three categories:
  1) **Low Risk** (if the model's output is 0 to 0.33)
  2) **Medium Risk** (if the model's output is between 0.34 and 0.66)
  3) **High Risk** (if the model's output is between 0.67 and 1)
- These risk levels are then displayed in the mobile app, providing users with a clear and comprehensible seven-day forecast of shark attack risk for their chosen location.
- It is important to know the risk forecasts rely on the accuracy of the marine weather forecasts as the forecasted marine weather variables are used as input in the forecasting model.

3.3. Conclusion

The mobile application serves as an intuitive bridge between end-users and a sophisticated AI model. Through a streamlined process, users can quickly ascertain the shark attack risk for any location, enabling them to make informed decisions about their marine activities. The backend, hosted on Google Cloud, ensures that the app remains responsive and accurate, utilizing real-time marine weather data and the power of deep learning to provide reliable forecasts.

4. Results

4.1. Introduction

In the domain of shark attack forecasting, the primary objective is to predict attacks accurately, ensuring the safety of individuals in and around marine waters.

Evaluating the efficacy of our model is, therefore, crucial in validating its applicability and usefulness. This section will present the results from our deep learning model, emphasizing its performance metrics, and comparing its predictions to existing preventative measures.

4.2. Model's Performance Metrics

- **Accuracy**:
- The model's forecast accuracy for the positive class on the test set is an impressive **0.8289**. This implies that in about 82.89% of instances, the model correctly predicted days with a higher risk of shark attacks.
- **Precision**:
- A precision score of **1.00** indicates that every time our model predicted a high-risk day, it was correct. In other words, there were no false positives.
- **Recall**:
- The model achieved a recall of **0.80**, suggesting that it was able to correctly identify 80% of actual high-risk days.
- **F1-Score**:
- With an F1-score of **0.89**, the model showcases a harmonious balance between precision and recall, ensuring that the model is neither too conservative nor too liberal in its predictions.
- **Confusion Matrix Visualization: (Figure 1)**
- For a more granular view of the model's performance, a confusion matrix has been included. This matrix provides a visual representation of the model's true positive, false positive, true negative, and false negative predictions. By analyzing this matrix, stakeholders can gain deeper insights into the model's strengths and areas of potential improvement.

4.3. Comparison to Existing Preventative Measures

While our model utilizes advanced machine learning and real-time marine weather data to forecast shark attack risks, the traditional preventative measures have been rather generic and are based on common-sense practices. Some of these include:

- Avoiding shiny jewelry as it can resemble fish scales and attract sharks.
- Limiting excessive splashing which can lure sharks closer.
- Refraining from swimming during early mornings or late evenings.
- Avoiding swimming during cloudy days or in murky waters as it reduces visibility for sharks, increasing the chances of accidental bites.

While these practices are sound advice, they are broad and do not offer specific guidance for a particular day or location. Moreover, these practices lean towards caution, potentially discouraging people from enjoying marine activities even on days with low shark attack risks.

Our AI-driven approach offers several advantages over these traditional measures:
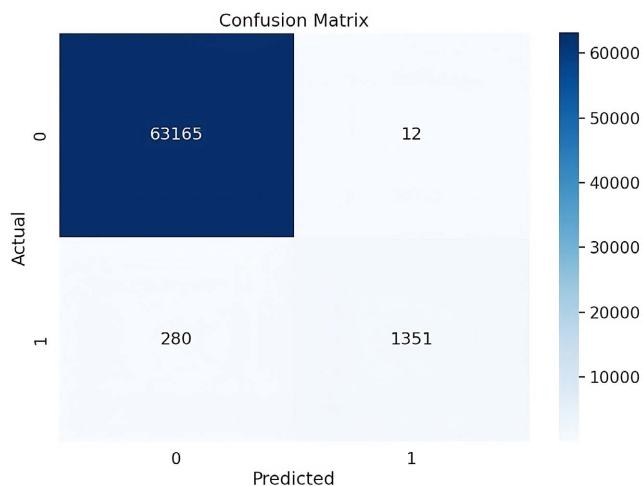
Figure 1. Confusion matrix visualization.

1) **Specificity**: By providing risk assessments for specific locations and days, individuals can make informed decisions about their activities.

2) **Real-time Data Integration**: The model's integration with marine weather data ensures that its predictions are based on current conditions, increasing its reliability.

3) **Accessibility**: Through a mobile app, users have instant access to risk assessments, making it more user-friendly than recalling a list of general best practices.

4.4. Conclusion

The results from our deep learning model are promising, showcasing high accuracy and precision. While traditional shark attack prevention advice is valuable, the integration of AI and real-time weather data offers a more dynamic, specific, and user-friendly approach to risk assessment. As the world continues to advance technologically, such tools can set a precedent, ensuring marine safety through data-driven insights.

5. Discussion

5.1. Introduction

The development and deployment of a predictive model for shark attack forecasting marks a pivotal moment in marine safety. As with all pioneering ventures, it's essential to evaluate its overall significance, understand how it compares to traditional methods, and determine areas that can be fine-tuned or expanded upon for future iterations. This section delves into the significance of our model, its improvements over existing measures, and potential areas of refinement.

5.2. Improvements over Existing Methods

- **Dynamic Update of Attack Data:**

- One of the most significant advantages of our model over traditional methods is its adaptability. As new shark attack data becomes available, the model can easily integrate this information to refine its predictive capabilities. Traditional methods remain static, but our model evolves, ensuring that it remains relevant and accurate over time.
- **Variable Inclusion and Expansion**:
- The ability to incorporate additional variables as they become accessible or as more research emerges means that our model can continuously grow in complexity and precision. This dynamic nature contrasts with the fixed guidelines of existing shark attack prevention advice, which can't be easily updated or expanded without broad public re-education campaigns.
- **Personalized and Location-specific Risk Assessment**:
- By offering location-specific risk assessments, our model provides more actionable insights than broad preventative guidelines. This tailored advice can empower individuals to make informed decisions about their safety.

5.3. Limitations and Areas for Improvement

- **Data Reliability and Completeness**:
- The efficacy of the model is intrinsically tied to the quality and completeness of the data it ingests. Any inaccuracies or gaps in the shark attack database could potentially skew predictions. Future iterations could focus on data verification mechanisms or diversifying data sources to create a more holistic dataset.
- **Generalization across Different Coastal Ecosystems**:
- Different coastal areas might have distinct marine ecosystems, which could influence shark behavior in unique ways. Tailoring the model to account for these regional differences could improve its predictive accuracy across various geographies.
- **Feedback Mechanism**:
- Currently, once a prediction is made, there is no feedback loop to confirm if the forecasted risk materialized or not. Integrating a user feedback mechanism within the mobile app could provide valuable real-world validation data.
- **Model Interpretability**:
- While neural networks are powerful predictors, they are often termed as "black boxes" due to their lack of interpretability. Ensuring stakeholders understand the model's decisions might be crucial for broader acceptance. Future work could focus on model transparency or incorporating explainable AI techniques.
- **External Factors**:
- Certain external factors, such as sudden changes in local fish populations, human activities like fishing tournaments, or marine celebrations, could influence shark movement and behavior. Including such events as additional variables could refine the model's predictive power.

5.4. Conclusion

The launch of our AI-driven shark attack forecasting model represents a substantial leap forward in marine safety. Its dynamic nature, ability to incorporate new variables, and location-specific predictions offer unparalleled advantages over traditional preventative measures. However, as with all innovative solutions, it's essential to acknowledge its limitations and continuously seek avenues for refinement. Embracing a mindset of continuous improvement ensures that the model remains at the forefront of marine safety innovations.

6. Project Conclusion

The journey from conceptualizing the need for an AI-driven shark attack forecasting model to its eventual deployment has been both challenging and enlightening. This research sought to harness the power of machine learning to fill a gap in marine safety—forecasting the risk of shark attacks. The importance of this endeavor was underpinned by data showing the alarming rise of such incidents, particularly in the U.S.

6.1 Main Findings

1) **Feasibility**: Through data collection, preprocessing, and model selection, it was established that it is indeed feasible to use available data to forecast future shark attack risks. With a focus on binary output that could be further classified into risk categories, the model displayed a promising ability to predict potential threats.

2) **Model Performance**: The fully connected neural network model demonstrated significant accuracy in its predictions, especially for the positive class. With a forecast accuracy of 0.8289 for positive instances, the model also showcased commendable precision, recall, and F1-score metrics. This signifies a robust prediction capability, especially given the novel nature of this venture.

3) **Comparison with Existing Methods**: Compared to traditional measures which revolve around broad guidelines, the model stands out with its dynamic, personalized, and location-specific risk assessments. While there's no direct benchmark in the realm of shark attack prediction, the model's approach surpasses general prevention advice both in granularity and adaptability.

4) **Implementation Insights**: The creation of a mobile app offers a direct, user-friendly interface to the public. By allowing users to input their locations and subsequently providing a tailored risk assessment, the application not only serves as a testament to the practical application of AI but also as an invaluable tool for marine safety.

6.2. Future Directions

1) **Data Expansion and Refinement**: As more shark attack data becomes available, future iterations of the model can integrate this information, refining its predictive power, improving accuracy. Furthermore, collaboration with ma-

rine biologists and shark experts can help in identifying new relevant data points.

2) **Model Variants**: Exploring other machine learning or deep learning architectures might offer even better predictive performance. Variations of neural networks, such as convolutional or recurrent architectures, might be worth investigating, especially as the dataset grows in complexity.

3) **User Feedback Integration**: By integrating a feedback mechanism within the mobile app, it's possible to collect real-world validation data. This feedback can be invaluable in continuously refining the model's accuracy.

4) **Broadening the Application**: While the current focus is on shark attacks, similar frameworks could potentially be applied to predict other marine-related incidents, setting a precedent for diverse use-cases.

In wrapping up, the successful development and deployment of the AI-driven shark attack forecasting model showcases the immense potential of artificial intelligence in pioneering new safety solutions. As the tides of technology continuously ebb and flow, this research stands as a beacon, illuminating the path to a safer marine future.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] International Shark Attack File (2023) Yearly Worldwide Shark Attack Summary. Florida Museum.
https://www.floridamuseum.ufl.edu/shark-attacks/yearly-worldwide-summary/#:~:text=Global%20total%20of%20unprovoked%20shark,average%20of%2072%20incidents%20annually

[2] Quartz Staff (2019) Shark Attacks Are on the Rise in the US and Australia. Quartz.
https://qz.com/1560529/shark-attacks-are-on-the-rise-in-the-us-and-australia/

[3] Stanich, A. (2023) Discover How Many Sharks Are Killed per Year and How You Can Help Them. A-Z Animals.
https://a-z-animals.com/blog/discover-how-many-sharks-are-killed-per-year-and-how-you-can-help-them/